

Research Paper ■

Computer Decision Support as a Source of Interpretation Error:
The Case of Electrocardiograms

THEODORE L. TSAI, MD, DOUGLAS B. FRIDSMA, MD, GUIDO GATTI, MA

Abstract Objective: The aim of this study was to determine the effect that the computer interpretation (CI) of electrocardiograms (EKGs) has on the accuracy of resident (noncardiologist) physicians reading EKGs.

Design: A randomized, controlled trial was conducted in a laboratory setting from February through June 2001, using a two-period crossover design with matched pairs of subjects randomly assigned to sequencing groups.

Measurements: Subjects' interpretive accuracy of discrete, cardiologist-determined EKG findings were measured as judged by a board-certified internist.

Results: Without the CI, subjects interpreted 48.9% (95% confidence interval, 45.0% to 52.8%) of the findings correctly. With the CI, subjects interpreted 55.4% (51.9% to 58.9%) correctly ($p < 0.0001$). When the CIs that agreed with the gold standard (*Correct CIs*) were not included, 53.1% (47.7% to 58.5%) of the findings were interpreted correctly. When the correct CI was included, accuracy increased to 68.1% (63.2% to 72.7%; $p < 0.0001$). When computer advice that did not agree with the gold standard (*Incorrect CI*) was not provided to the subjects, 56.7% (48.5% to 64.5%) of findings were interpreted correctly. Accuracy dropped to 48.3% (40.4% to 56.4%) when the incorrect computer advice was provided ($p = 0.131$). Subjects erroneously agreed with the incorrect CI more often when it was presented with the EKG 67.7% (57.2% to 76.7%) than when it was not 34.6% (23.8% to 47.3%; $p < 0.0001$).

Conclusions: Computer decision support systems can generally improve the interpretive accuracy of internal medicine residents in reading EKGs. However, subjects were influenced significantly by incorrect advice, which tempers the overall usefulness of computer-generated advice in this and perhaps other areas.

■ J Am Med Inform Assoc. 2003;10:478–483. DOI 10.1197/jamia.M1279.

The Institute of Medicine's report, *To Err Is Human: Building a Safer Health System*, has resulted in an intense effort to use information technology as a means to reduce medical errors. The report suggested that decision support systems, particularly provider order entry, are an important component of reducing medical errors and costs.¹ One aspect of this belief is that an expert system will give advice to the provider at the

point of order entry and that the provider will appropriately accept or reject the advice. Computer-based expert systems are a type of decision support system meant to improve health care quality through the advice given in the form of alerts, reminders, or summaries.² Whereas previous work has suggested that computer-based decision support shows great promise in improving health care by enhancing reasoning and decision making by physicians,² there have been no studies regarding the possible negative effects.

The most widespread application of an expert system in clinical care is the electrocardiogram (EKG) expert system.³ Previous work supports the notion that the accuracy of the EKG expert system approaches the accuracy of cardiologists,⁴ and most researchers believe that EKG expert systems are therefore helpful to the physician.^{3,5,6} Laks and Selvester⁷ state that in their experience, "physicians aided by computers produce the best interpretation of EKGs." However, few studies have quantified this effect, and none have addressed the effect that incorrect CI has on physician interpretive accuracy.

A common opinion regarding the CI of EKGs is that even if the interpretation is not correct, it still provides useful information. Macfarlane⁵ suggests that the CI "at the very least

Affiliation of the authors: Center for Biomedical Informatics and Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania.

This work was supported by training grant 5T15 LM/DE07059 from the National Library of Medicine. The authors thank the following individuals for their assistance: Edward Curtiss, MD, for providing a second cardiologist interpretation of the EKG study cases; Charles Tsai, MD, for providing the EKG study tracings and expert assistance with categorizing the computer findings; Charles Friedman, PhD, and Cynthia Gadd, PhD, for assistance with study structure; and the internal medicine residents of the University of Pittsburgh.

Correspondence and reprints: Theodore L. Tsai, MD, Center for Biomedical Informatics and Department of Medicine, University of Pittsburgh, Suite 8084 Forbes Tower, 3600 Meyran Avenue, Pittsburgh, PA 15261; e-mail: <tsai@cbmi.upmc.edu>.

Received for publication: 10/23/02; accepted for publication: 05/05/03.

provides a second opinion which can be accepted or rejected by a physician." Whether this second opinion is always a benefit to the physician is unproven; although two studies have examined how accurate physicians are in accepting correct computer generated advice,^{3,6} no studies have examined how physicians use incorrect computer advice.

The central question guiding our investigation was to what extent does a "correct" CI relate to a correct interpretation and to what extent does an "incorrect" CI relate to an incorrect human interpretation over a set of typical tracings by noncardiologist physicians? One study used cardiologists as subjects who were experts in the area of EKG interpretation and showed that cardiologists are both precise and accurate in EKG interpretation with and without computer assistance.³ Internists, however, are much less precise and accurate than cardiologists in detecting and correcting computer errors.⁸ We chose to use internal medicine residents as our subject group because they are not experts and usually are the first physicians to interpret EKGs in academic hospitals.

Methods

We used a two-period crossover design with matched pairs of subjects randomly assigned to sequencing groups to obtain clinicians' interpretations of two equally difficult sets of EKGs (labeled set A and set B).⁹ To evaluate the effect of the CI, subjects were matched by year of postgraduate training and divided into two groups (designated AB and BA). Group AB first interpreted set A without CI support and then interpreted set B with computer support. Group BA first interpreted set B without CI support and then interpreted set A with computer support. No clinical information was given because different clinical histories have been shown to affect the physicians' interpretation of identical EKGs.^{10,11} The effect of the CI on subject accuracy was determined by comparing the percentage of findings each subject correctly interpreted without the CI with the percentage of findings correctly interpreted with the CI. This effect also was broken down by categories based on the computer's advice (i.e., *Correct*, *Incorrect*, *Nonspecific*, or *Wrong by Exclusion*).

Study Site

The study was based at an academic medical center, the University of Pittsburgh Medical Center. The principal investigator, first cardiologist, and general internal medicine collaborator were based at the University of Pittsburgh. The second cardiologist collaborator was based in private practice at the Gould Medical Foundation in Modesto, California.

Case Materials

The second cardiologist provided 83 EKGs from actual episodes of care from his residency, cardiology fellowship training, and private practice. From this pool, the principal investigator and second cardiologist examined the EKGs for

three factors: (1) the EKG was accompanied by an interpretation by a cardiologist, (2) the EKG was accompanied by CI, (3) the EKG was of good reproductive quality and without extraneous markings. Eight tracings did not meet the above criteria and were discarded. The remaining 75 EKGs then were selected for the presence of the basic findings that physicians at the end of their internal medicine residency should be able to identify, using the list developed by Pinkerton and his group as a guide.¹² In so doing, we eliminated findings that would require an expert cardiologist to interpret and were left with findings that residents after their first year of training would be expected to interpret correctly.

Identifying information and the CI were removed from these EKGs, and the tracings were presented to the first cardiologist for a second interpretation. The interpretations of the two cardiologists were compared, and the EKGs in which there was disagreement between the two cardiologists were discarded. By doing this, we validated a set of EKGs that could be interpreted by the tracing alone and that there was no degradation of the tracings in the copying process. This was a similar method to that used in a previous study in establishing a gold standard for a set of EKG interpretations.⁸

Categorization of the Findings

From the remaining tracings, the principal investigator and second cardiologist compared the gold-standard cardiologists' findings with the computer reports to place each finding into one of four categories (Table 1): (A) *Correct*, the correct finding was given by the CI, (B) *Nonspecific*, the CI mentioned the abnormality on the tracing but did not give the diagnosis, (C) *Wrong by exclusion*, the CI did not mention the abnormality at all, (D) *Incorrect*, the CI of the finding was incorrect in its final diagnosis.

Willems et al.⁴ estimated that the CI has an accuracy rate that ranges from 42% to 96%. The principal investigator and second cardiologist selected a group of EKGs in which the computer interpretations were correct approximately 60% of the time, the usual accuracy of the CI. We also wanted to mimic the prevalence of the findings in the real world. For example, the most common rhythm encountered in practice is normal sinus rhythm, and this was the most common rhythm that was in the test set, appearing 11 times. Atrial fibrillation is a common rhythm in practice but is much less common than normal sinus, and this is also reflected in the test set. After this selection process, our test set consisted of 23 EKGs with 54 total findings, 32 of which were correctly (59.2%) interpreted by the CI. Table 2 lists the findings and the number of times they appeared in the test set. Notably, 15 findings were presented only one time. "No left atrial enlargement" was included as a finding because on one

Table 1 ■ An Example of the Categorization of the Computer Interpretation Compared with the Gold Standard Cardiologists' Interpretation

Cardiologists' Interpretation	Computer Interpretation (CI)	Category of CI
Normal sinus rhythm	Normal sinus rhythm	Category A—Correct
Left anterior fascicular block	Left axis deviation	Category B—Nonspecific
Left atrial enlargement	[Does not mention]	Category C—Wrong by exclusion
Early ventricular repolarization	Widespread ST elevation, consider pericarditis	Category D—Incorrect

Table 2 ■ EKG Findings and the Number of Times They Appear in the Test Set

Atrial fibrillation (2)
Anterior/septal MI (2)
Inferior wall MI (3)
ST-T changes consistent with digitalis/digoxin/digitoxin effect (2)
Normal sinus rhythm (11)
Sinus bradycardia (1)
Prolonged QT (1)
Left atrial enlargement (2)
Second-degree AV block Mobitz type I (1)
First-degree AV block (1)
Insignificant ST elevation, no MI (3)
NSR with 2:1 AV block (1)
No left atrial enlargement (1)
Left bundle branch block (1)
Incomplete right bundle branch block (1)
Right bundle branch block (3)
Left anterior fascicular block (2)
Left posterior fascicular block (1)
Sinus arrhythmia (1)
Multifocal atrial tachycardia (2)
Normal sinus rhythm with frequent premature atrial contractions (1)
Electrode reversal (1)
ST elevation consistent with early ventricular repolarization (2)
Left ventricular hypertrophy (3)
Atrial flutter (2)
Supraventricular tachycardia (1)
Nonspecific ST-T changes (1)
Ventricular tachycardia (1)

occasion the CI listed it as a finding when it was not actually present. The EKGs then were divided into two sets, taking care to keep the variety and correctness equal between the two (Table 3).

Subjects

We recruited 30 internal medicine residents who were either in their second or third years of training. Subjects were enrolled on a rolling basis. All subjects graduated from accredited allopathic American medical schools. No compensation was given for participation in the study. Data were collected over a five-month period from February through June 2001.

Procedure

Subjects were stratified by experience level and then assigned randomly to either group AB or BA. Assignment was balanced such that experience levels were distributed equally between the two groups with six second-year residents and nine third-year residents in each group.

Subjects were presented with the EKG tracings and were asked to record their interpretations on a blank sheet of paper

Table 3 ■ Categorization of the Computer Interpretations of Each Set of EKGs

Set A	Set B
11 Tracings	12 Tracings
27 Findings	27 Findings
16 Correct	16 Correct
3 Nonspecific	1 Nonspecific
2 Wrong by exclusion	4 Wrong by exclusion
6 Incorrect	6 Incorrect

as they would normally do in a patient's medical record. Group AB first interpreted EKG set A without CI support and then set B with computer support. Group BA first interpreted set B without CI support and then set A with computer support. Subjects were instructed to interpret the tracings in the order presented and not to return to previously read EKGs to change answers. All data collection sessions were proctored, and all subjects completed their interpretations within the 1-hour time limit. A board-certified internist, blinded to subject and whether CI accompanied the individual tracings, scored the subjects' stated interpretations as either *correct* (consistent with the gold-standard interpretations as determined by the two cardiologists) or *incorrect* (not consistent with the gold-standard interpretations). Scoring guidelines were used to maintain consistent interpretation of ambiguous answers (i.e., in cases in which hedging occurred or where several possible diagnoses were given for the same finding, the subject was scored as incorrect for that finding). When the subject's interpretation was incorrect, the internist also noted if the subject's interpretation was consistent with the CI. The protocol was approved as an exempted review by the University of Pittsburgh's Institutional Review Board.

Scoring Metrics

We assessed subjects' interpretive accuracy with a binary measure, which credited the subject with a correct interpretation if the correct interpretation appeared on the subject's answer sheet.

A second binary measure assessed the negative effects of incorrect CI on subjects' interpretations. For the subset of findings in which the CI was incorrect and the subject interpretation was incorrect, the subject response was noted to have agreed or disagreed with the CI.

Analysis

Three separate analyses provided complementary views of the results. The first used the binary outcome to measure subject's interpretation correctness to estimate the effect of CI on accuracy. The percentage of findings correctly interpreted with the CI was compared with the percentage correctly interpreted without the CI, treating the presence of CI as within EKG set within subjects. The second analysis also used the subject's correctness of interpretation outcome measure with category of finding (i.e., *Correct*, *Nonspecific*, *Wrong by Exclusion*, and *Incorrect*) and presence of the CI as independent variables. This analysis looked at the differences in the proportions of findings correctly identified by subjects for levels of CI within each category of computer interpretation. The third analysis sought to elucidate whether a CI that was *Incorrect* was a factor in subjects incorrectly interpreting findings. For the subset of findings in which the CI was *Incorrect* and the subject failed to correctly interpret the finding, the subject's response was scored as agreeing with the CI or not agreeing. If subject responses agreed with the CI more often when the CI was included, this would suggest that the CI had influenced the subjects negatively.

Although subjects are the unit of experimentation (i.e., randomly assigned to levels of treatment) and the primary sampling unit (i.e., independently sampled), the units of analysis are the subjects' findings, which are assumed to be

dependent and nested within subjects. Although findings are nested within EKG tracings, it is assumed that this grouping of findings has no effect on subject accuracy or the dependence of the findings. We used the Generalized Linear Model (GzLM) procedure¹³ for each analysis, assuming the binary outcome is distributed as a Bernoulli variable with a logit link and used naive empirical variance estimates¹⁴ for the model effects to account for the dependence of findings within subjects. Wald statistics were used to statistically test the observed results against the null condition at the $\alpha = 0.05$ level. Two-tailed 95% confidence intervals were calculated for percents by transforming logit scale Wald intervals (i.e., using empirical standard error estimates) into percent scale intervals.¹⁵ Confidence intervals were calculated for percent differences using the method described by Newcombe and Altman¹⁶ using the transformed empirical Wald limits in place of the Wilson limits.

Results

Percentage of Findings Correctly Interpreted by Subjects

The complete data set included 54 findings (27 with CI and 27 without CI) and 30 subjects, for a total of 1,620 subject-finding data points. Without the CI, subjects correctly interpreted an average of 13.2 of 27 findings correctly (48.9%; 95% confidence interval; 45.0% to 52.8%), which increased to 15.0 of 27 (55.4%; 51.9% to 58.9%) with CI support, a change of 6.6% (1.3% to 11.8%; $p < 0.0001$). This suggests that the CI has an overall positive influence on resident physicians' accuracy in interpreting the EKGs in this set.

Correctness of the CI

The goal of the next analysis was to evaluate how the correctness of the CI affects the likelihood that the finding will be interpreted correctly. Table 4 summarizes by category the percentage of findings that were correctly interpreted.

Correct

The CI was *Correct* in 32 of 54 findings. Without the CI, 255 of 480 (53.1%; 47.7% to 58.5%) subject findings were correctly interpreted; this increased to 327 of 480 (68.1%; 63.2% to 72.7%) with CI, a 15% difference (7.7% to 22.1%; $p < 0.0001$), suggesting that the CI, when *Correct*, increased the likelihood that the finding would be correctly interpreted.

Incorrect

The CI was *Incorrect* in 12 of 54 findings. Without the CI, 102 of 180 (56.7%; 48.5% to 64.5%) subject findings were interpreted correctly; this decreased to 87 of 180 (48.3%; 40.4% to 56.4%) subject findings when the CI was included

(−8.4%; −19.5% to 3.1%; $p = 0.13$). The trend toward the decrease in the likelihood that the findings would be correctly interpreted did not reach statistical significance.

Nonspecific and Wrong by Exclusion

Only four of 54 and six of 54 findings were *Nonspecific* and *Wrong by Exclusion*, respectively. For the *Nonspecific* category, 20 of 60 (33.3%; 24.5% to 43.5%) subject findings in which the CI was not included were interpreted correctly; 15 of 60 (25.0%; 16.6% to 35.9%) subject findings were interpreted correctly with the CI (−8.3%; −21.5% to 5.7%; $p = 0.22$).

In the *Wrong by Exclusion* category, 19 of 90 (21.1%; 13.6% to 31.2%) subject findings in which the CI was not shown were interpreted correctly; 20 of 90 (22.2%; 15.3% to 31.1%) subject findings were interpreted correctly with the CI (1.1%; −11.1% to 12.7%; $p = 0.85$). The results do not approach significance in these two categories.

Analysis to Elucidate the Effect of Incorrect CI on Subjects

The CI was categorized as *Incorrect* in 12 of 54 findings. When the incorrect CI was not presented in this subset of findings, subjects were incorrect in 78 of 180 subject findings. Even though subjects had no knowledge of what the computer report was, there was subject agreement with the incorrect CI in 27 of these 78 subject findings (34.6%; 23.8% to 47.3%). When the incorrect CI was presented to subjects, the incorrectness rate increased to 93 of 180 subject findings. There was agreement with the incorrect CI in 63 of 93 findings (67.8%; 57.2% to 76.7%). The results are summarized in Table 5. This difference is statistically significant (33.1%; 16.6% to 47.2%; $p < 0.0001$), implying that the CI specifically was influential in misleading subjects. Subjects were prone to agree with the CI, even though it was incorrect.

Discussion

The reasons for the EKG expert system's acceptance in clinical care is based on its perceived accuracy, ease of use, low cost, and the assignment of the responsibility of final interpretation to the physician.^{5,17} In this study, we have shown that the impact of computer assistance on nonexpert subject performance in interpreting EKGs depends on the correctness of the advice given. More specifically, we have examined the situation in which the CI is incorrect and illustrated how it can be a detrimental influence on decision making. A common opinion regarding the CI in clinical practice is that, at the very least, the physician with the CI is no worse at interpreting EKGs than the physician reading the EKG without computer support.⁵ The results of our study suggest otherwise.

Table 4 ■ Likelihood That a Finding Would Be Interpreted Correctly Depending on Presence of the Computer Interpretation

	Computer Interpretation Correct ($n = 32$) Category A	Computer Interpretation Nonspecific ($n = 4$) Category B	Computer Interpretation Wrong by Exclusion ($n = 6$) Category C	Computer Interpretation Incorrect ($n = 12$) Category D
No CI	53.1%; (47.7%, 58.5%)	33.3% (25.5%, 43.5%)	21.1%; (13.6%, 31.2%)	56.7%; (48.5%, 64.5%)
With CI	68.1%; (63.2%, 72.7%)	25.0%; (16.6%, 35.9%)	22.2%; (15.3%, 31.1%)	48.3%; (40.4%, 56.4%)
Difference	+15.0% (7.7%, 22.1%)	−8.3% (−21.5%, 5.7%)	+1.1% (−11.1%, 12.7%)	−8% (−19.5%, 3.1%)
p-value	< 0.0001	0.222	0.854	0.131

Note. Grouped by computer interpretation category. 95% confidence intervals are given.

Table 5 ■ Proportion of Subject Findings in Which the CI Was Incorrect that Were Incorrectly Interpreted by the Subject and Agreed with the Incorrect Computer Interpretation

No CI	34.6% (23.8%, 47.3%)
With CI	67.7% (57.2%, 76.7%)
Difference	−33.1% (16.6%, 47.2%)
p-value	< 0.0001

There was a clear and significant improvement in the percentage of findings interpreted correctly when the subjects received the CI; this result is similar to what has been shown in previous work.⁶ However, when the findings are isolated by correctness, we see that there is a striking difference between how correct advice influences subjects and how incorrect advice influences the same subjects. There is a dramatic improvement in subject performance when the CI is correct and is provided to the subjects, but there is a trend that subject accuracy decreases when the CI is incorrect. Sample size might explain why this detrimental effect did not quite reach significance: the category in which the CI was *Incorrect* had 360 physician-finding data points compared with 960 in which the CI was *Correct*. Nevertheless, this trend goes against the belief that the CI is at least not detrimental in physician interpretation of EKGs.

In fact, the CI appears to be the cause of subjects' poor performance. Table 5 shows that in the subset of findings in which the CI was incorrect, subjects wrote an interpretation that agreed with the incorrect CI almost twice as often when they were assisted by the computer than when they were not (67.7% vs. 34.6%). Subjects had a clear tendency to agree with the CI even though it was *Incorrect*.

The degree of difficulty each EKG tracing presents to the interpreter is subjective and difficult to quantify. One may believe that the findings in which the CI was *Incorrect* would be inherently more difficult to interpret and therefore more often interpreted incorrectly by subjects when the CI was not present compared with the findings in which the CI was *Correct*. Notably, in the findings in which the CI was *Incorrect* but not provided, findings were interpreted correctly 56.7% of the time versus 53.1% of the findings in which the CI was deemed *Correct* but was not provided (Table 4). If the findings in which the CI was *Correct* were easier to interpret than those in which the CI was *Incorrect*, subjects without the CI influencing their interpretation should have agreed with the cardiologists in a higher percentage of findings in which the CI was *Correct*.

A recent study that examined the effect of computer-generated EKG reports on emergency senior houseofficers showed no significant difference in the EKG interpretation error rate when junior doctors had assistance from computer-generated reports.¹⁸ The authors suggested that "practice might be improved by instructing [resident physicians] to seek senior advice whenever they disagree with, or do not understand, a report." Although it is reasonable to seek expert advice if a resident physician does not agree with a CI, it is not necessarily reasonable for a resident physician to believe that his or her interpretation is correct if they agree with the CI. Our results suggest that the bias inherent in

reading a CI does not allow the resident, nonexpert physician to accurately judge when the CI is incorrect.

Another study investigated whether the CI guides or misleads the EKG interpretation by physicians in the emergency room.¹⁹ The authors concluded that "Whether an incorrect computer diagnosis was provided or not, did not significantly influence the physicians' conclusions," and "Computer-based ECG diagnoses seem to be helpful to emergency ward physicians, but a certain level of ECG experience is required to utilize the program." Although we agree that experience with EKG interpretation will improve the interpreter's ability to use the computer's advice (cardiologists have been shown to be precise and accurate at accepting and rejecting computer-based EKG diagnoses³), we feel that the study design lacked the power to determine how incorrect computer diagnoses influence physicians. Their study consisted of 20 subjects who interpreted ten EKGs that randomly had the computer diagnosis accompanying it (which were either all correct or all incorrect when present) versus our study's 30 subjects who each read the same set of 23 EKGs, which included both correct and incorrect computer diagnoses when present. Because we did not reach statistical significance in how incorrect diagnoses affect physician interpretation, it is not surprising that their smaller study size also did not reach significance and possibly was too small to detect a trend. Most notably, the authors also did not take the next step to determine whether the physicians who misinterpreted an EKG did so *because of* the incorrect computer diagnosis. There is no comment on agreement with incorrect computer diagnoses, which is a major focus of our study.

The goal of this study was to question the popular belief that physicians are good at incorporating good information while at the same time rejecting bad information in the realm of computer-interpreted EKGs, and we purposely did not address clinical outcomes. Although determining clinical significance would be the ultimate goal, the only published literature to support this belief were editorials or opinions based on observation^{5,7}; we found no research to support this idea, and we first had to show that the belief that the CI did at least not cause harm is a faulty notion at its root—that, in fact, physicians are significantly negatively influenced by incorrect computer interpretations of EKGs. If we were to consider clinical importance in this study, we would have had to include clinical history, which has proven to alter physician interpretations of EKGs. Hatala et al.^{10,11} have described how the interpretation of EKGs can change when different clinical history is given (and also when no clinical history is given). The question of exactly how much historical information to provide adds another layer of complexity. By eliminating clinical history, we distilled the EKG interpretation task to its purest form. Extrapolating the EKG findings to clinical significance involves another level of decision making that is beyond the scope of this study.

Further work is needed to elucidate the reasons for the subjects' poor performance when receiving the CI when it was incorrect. One paradigm represents the EKG expert system as a tool and whether targeted instruction on the use of this tool would improve interpretive accuracy. Cardiologists have been shown to be more skilled than internists at using the CI to interpret EKGs accurately.⁸ If

cardiologists acquire this skill by specific instruction rather than volume of experience, it might be possible to develop a program aimed at noncardiologist physicians to instruct them on the proper use of the EKG expert system to improve their use of the CI.

The true reason is likely a combination of training, experience, and other human cognitive factors. There may be a subconscious bias toward believing a computer-generated result like the CI or, similarly, a lack of faith in subjects' own EKG interpretation skills. We also cannot ignore the effects of anchoring bias, as described by Tversky and colleagues²⁰:

"In many situations, people make estimates by starting from an initial value that is adjusted to yield the final answer. The initial value, or starting point, may be suggested by the formulation of the problem, or it may be the result of a partial computation. In either case, adjustments are typically insufficient."

In terms of anchoring, the initial value in the case of the EKG is the CI. The results in the section "Nonspecific and wrong by exclusion" suggest that subjects are unable to sufficiently adjust from the initial value as Tversky and colleagues described.

Conclusions

This study is a first step in understanding the divide between computer decision support and human decision making. Even though the EKG expert system's performance has been comparable with that of gold-standard cardiologists,⁴ the effects of the system on its intended noncardiologist physician end users had not been addressed. We have shown the possible negative side to its use. Whether the detrimental influence on EKG interpretation in this study would have resulted in patient harm or other undesirable outcomes is yet to be determined and, for the purposes of this study, not essential. To determine if medical error could result from bad advice, we had to first show that bad advice cannot be as easily dismissed as has been thought.

The acceptance of any decision support system is based on a number of factors, and it seems that the human cognitive factors are the least of them, being superseded by cost and ease of use. It is important in this and future systems to consider how the information is processed by the end user, because there may be unforeseen detrimental effects. Expert systems like the EKG system show promise to assist nonexperts in diagnosing and treating patients. However, we should temper our exuberance to implement these systems—further work is needed to examine the risks of incorrect advice on physician decision making before deploying them.

References ■

1. Kohn LT, Corrigan JM, Donaldson MS, (eds). *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press, 1999.
2. Friedman CP, Elstein AS, Wolf FM, et al. Enhancement of clinicians' diagnostic reasoning by computer-based consultation. *JAMA*. 1999;282:1851–6.
3. Brailer DJ, Kroch E, Pauly MV. The impact of computer-assisted test interpretation on physician decision making: the case of electrocardiograms. *Med Decis Making*. 1997;17:80–6.
4. Willems JL, Abreu-Lima C, Arnaud P, et al. The diagnostic performance of computer programs for the interpretation of electrocardiograms. *N Engl J Med*. 1991;325:1767–73.
5. Macfarlane PW. A brief history of computer-assisted electrocardiography. *Methods Inf Med*. 1999;29:1803–4.
6. Hillson SD, Connelly DP, Liu Y. The effects of computer-assisted electrocardiographic interpretation on physicians' diagnostic decisions. *Med Decis Making*. 1995;15:107–12.
7. Laks MM, Selvester RH. Computerized electrocardiography—an adjunct to the physician. *N Engl J Med*. 1991;325:1803–4.
8. Pryor WM, Blackhurst DW. Follow-up evaluation of computerized electrocardiographic interpretations: a comparison of cardiologists' vs. internists' performance. *J S C Med Assoc*. 1996;92:339–343.
9. Fleiss JL. *The Design and Analysis of Clinical Experiments*. New York, NY: John Wiley & Sons, 1986.
10. Hatala RA, Norman GR, Brooks LR. The effect of clinical history on physicians' ECG interpretation skills. *Acad Med*. 1996;(10 suppl):S68–70.
11. Hatala RA, Norman GR, Brooks LR. Impact of a clinical scenario on accuracy of electrocardiogram interpretation. *J Gen Intern Med*. 1999;14:126–9.
12. Pinkerton RE, Francis CK, Ljungquist KA, Howe GW. Electrocardiographic training in primary care residency programs. *JAMA*. 1981;246:148–50.
13. McCullagh P, Nelder JA. *Generalized Linear Models*, Second Edition. New York: Chapman & Hall, 1991.
14. Liang K, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73:13–22.
15. Neter J, Kutner MH, Nachtsheim CJ, Wasserman W. *Applied Linear Regression Models*, Third Edition. Chicago: Irwin, 1996, pp 603–5.
16. Newcombe RG, Altman DG. Proportions and their differences. In: Altman DG, Machin D, Bryant TN, Gardner MJ (eds). *Statistics with Confidence, Confidence Intervals and Statistical Guidelines* (ed 2). London, UK: BMJ Books, 2000.
17. Alter S. *Information Systems: A management Perspective* (ed 3). New York: Addison-Wesley, 1999.
18. Goodacre S, Webster A, Morris F. Do computer generated ECG reports improve interpretation by accident and emergency senior house officers? *Postgrad Med J*. 2001;77:455–7.
19. Morisbak B, Gjesdal K. [Computer-based interpretation of ECG—guiding or misleading?]. *Tidsskrift for Den Norske Laegeforening*. 1999;119:3441–4.
20. Kahneman D, Slovic P, Tversky A. *Judgement under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press, 1982.